

---

---



# NICTER Dataset 2021

---

---

笠間 貴弘

国立研究開発法人情報通信研究機構  
サイバーセキュリティ研究所 サイバーセキュリティ研究室  
研究マネージャー

# NICTER Dataset 2021

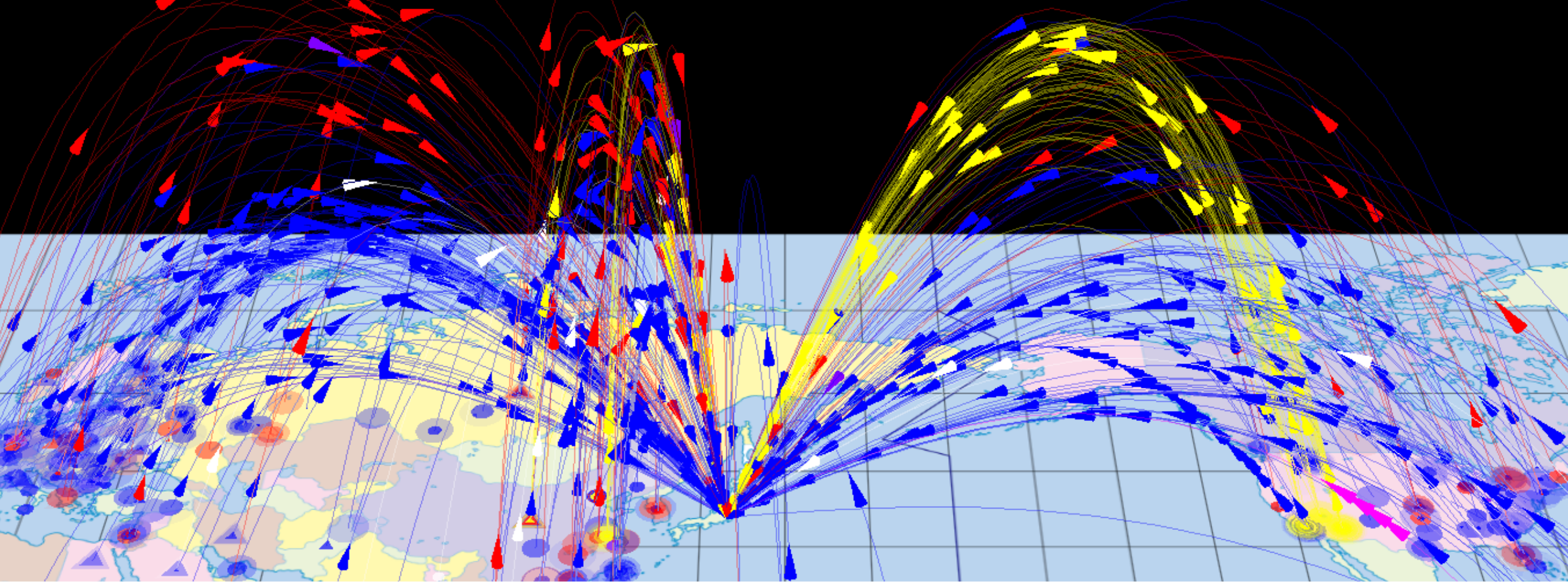
---

## ● ダークネットトラフィックデータ

- ✓ /20(約4千アドレス)のダークネットトラフィック
- ✓ 観測期間は2011年4月1日から現在まで**10年間以上**
- ✓ NONSTOP上で提供 (pcap+DB)

## ● スпамメールデータ

- ✓ NICTのメールサーバに届いたダブルバウンスメール
- ✓ 観測期間は2015年1月1日から現在まで6年間以上
- ✓ NONSTOP上で提供 (メールファイル)



# NICTER

- サイバー攻撃リアルタイム大規模観測・分析システム
- 国内外で30万の未使用IPアドレス“ダークネット”を観測
- 無差別型サイバー攻撃の大局的な傾向把握に有効

# ダークネット観測とは？

- **ダークネット：未使用のIPアドレス空間**

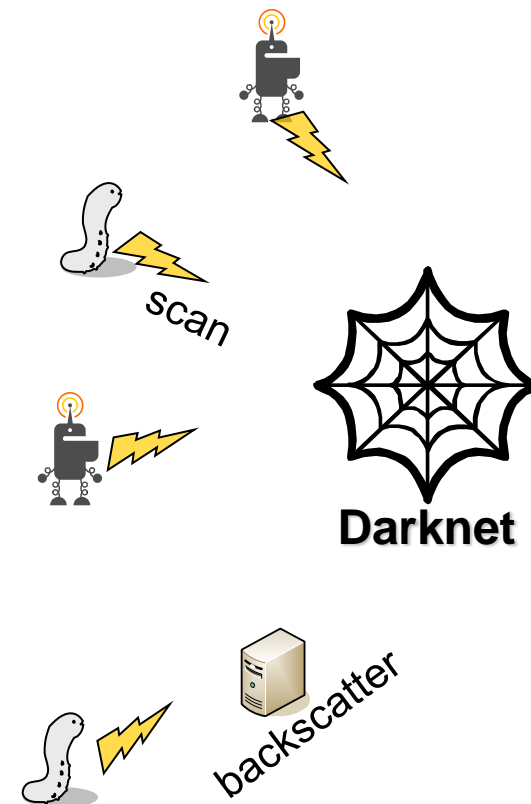
- ✓ 正常な通信は“基本的に”届かない

- **実際は大量の通信が届く**

- ✓ マルウェアによるスキャン
- ✓ DDoS攻撃の跳ね返り
- ✓ リフレクション攻撃の準備活動
- ✓ etc.

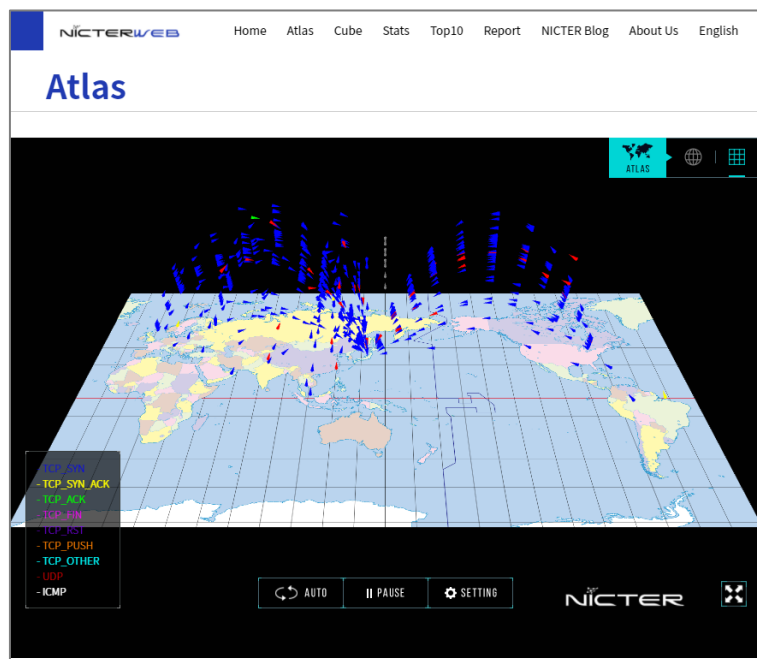
- **ダークネットの観測によって  
パンデミックの兆候が分かる**

- ✓ パンデミック：マルウェアの大量感染



# 観測結果や分析結果は一般公開もしています

- NICTERWEB (<http://www.nicter.jp/>)
- NICTER Blog (<http://blog.nicter.jp>)
- NICTER 観測レポート 2016~2020



The NICTER Blog page features a header with the NICTER logo and navigation links. The main content includes a post titled "サイバー脅威情報集約システム EXIST" (Cyber Threat Intelligence System EXIST) and another post titled "継続する 5555/TCP ポート宛攻撃通信と ADB が有効化された脆弱な Android エミュレータについて" (Continuing 5555/TCP Port-targeted Attack Communications and About Vulnerable Android Emulators with ADB Enabled).

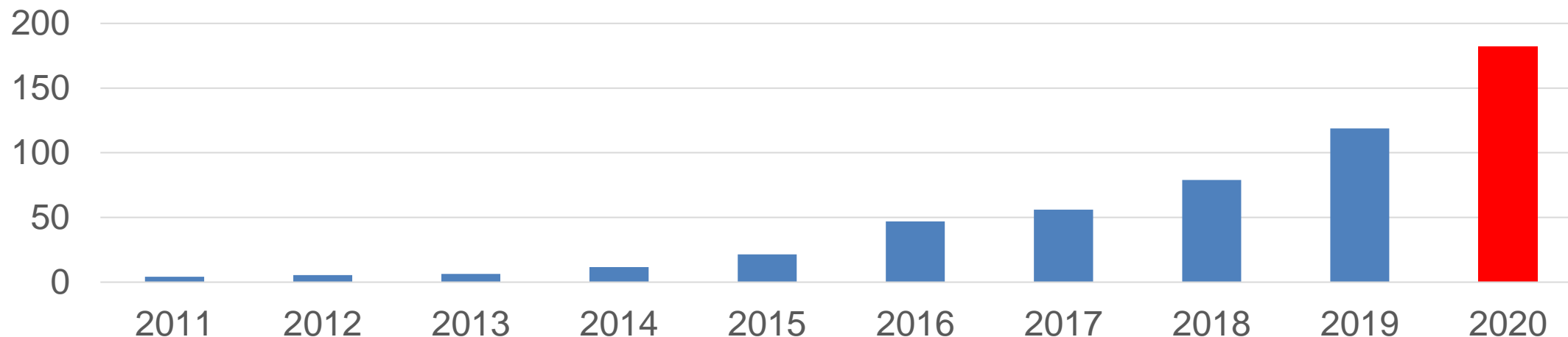
# よくある誤解（その1）

Q：ワームとかもう古いしスキャンなんて飛んでこないでしょ？

A：なんかそういうデータあるんですか？



(パケット数, 単位: 万)

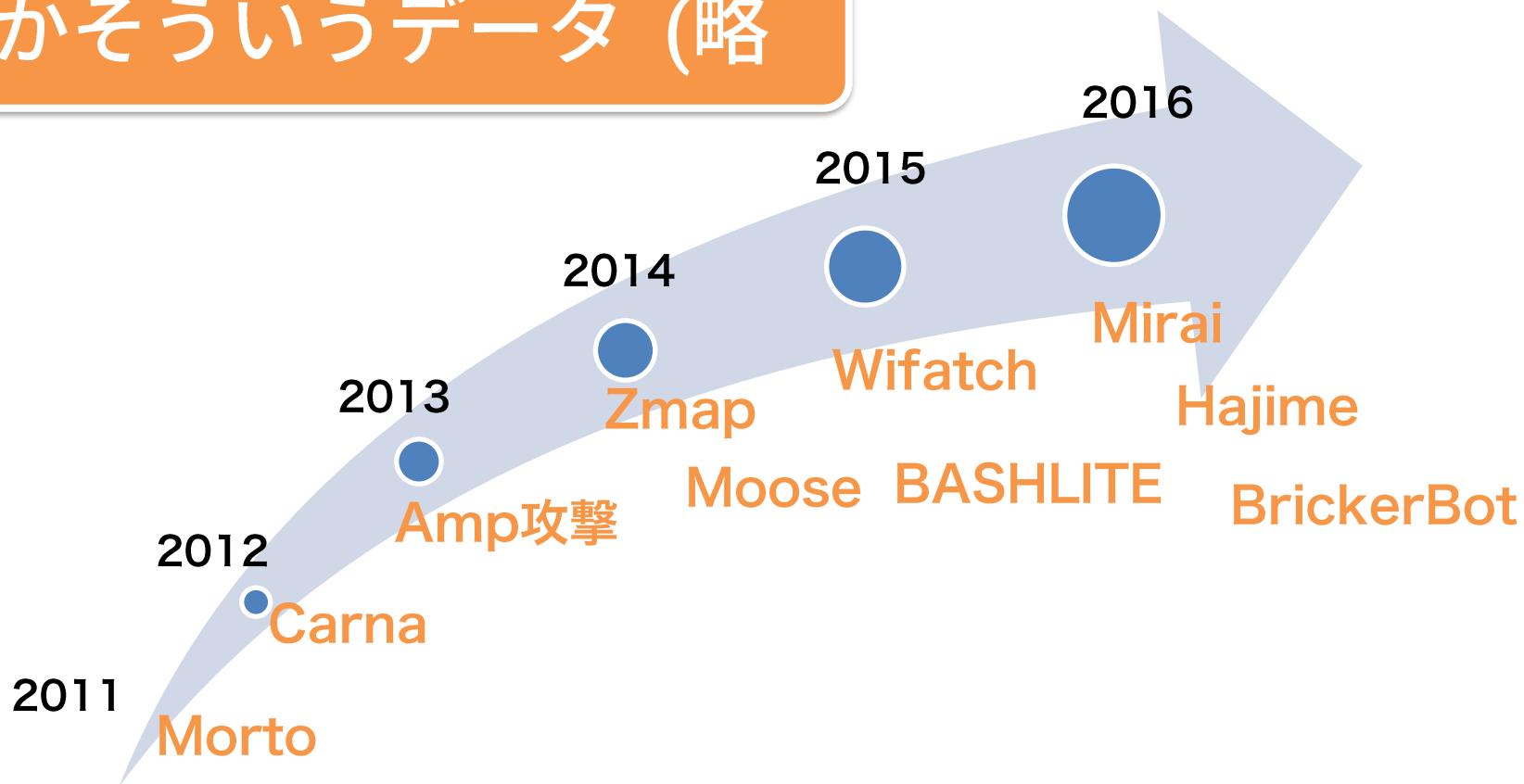


1 IPアドレスあたりの年間総観測パケット数

# よくある誤解（その2）

Q：スキャンしているのなんて古いConfickerとかだけじゃないの？

A：なんかそういうデータ（略



# よくある誤解（その3）

Q：今更ダークネットデータ使った論文とか出てないんじゃない？

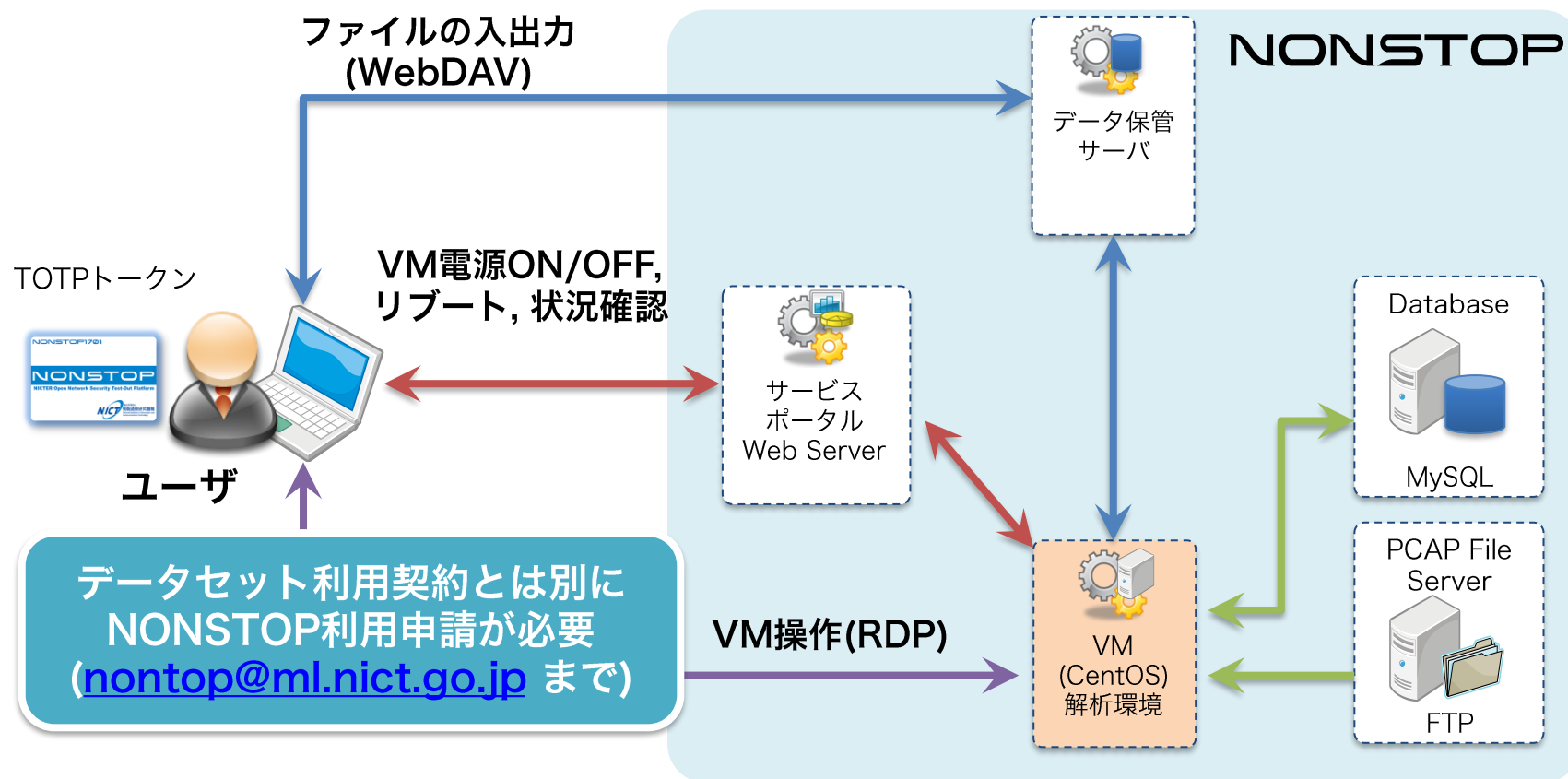
A: な (略)

<p>This paper appeared in <i>Proceedings of the 23rd USENIX Security Symposium</i>, August 2014.</p> <p>An Internet-Wide View of Internet-Wide Scanning</p> <p><b>USENIX Sec '14</b></p> <p><b>Abstract</b></p> <p>While it is widely known that port scanning is widespread, neither the scanning landscape nor the defensive reactions of network operators have been measured at Internet scale.</p> <p>scanning, and successfully fingerprint ZMap and Masscan. We present a broad view of the current scanning landscape, including analyzing who is performing large scans, what protocols they target, and what software and</p>	<p>Amplification Hell: Revisiting Network Protocols for DDoS Abuse</p> <p><b>NDSS '14</b></p> <p><b>Abstract</b></p> <p>In distributed attacks, adversaries send requests to recursive DNS resolvers and spoof the IP address of a victim. These servers, in turn, flood the victim with valid responses and – unknowingly – exhaust its bandwidth. Recently, attackers launched DDoS attacks with hundreds of C&amp;A bandwidth of this kind. While the attack technique is well-known for a few protocols such as DNS, it is unclear if further protocols are vulnerable to similar or worse attacks.</p> <p>In this paper, we revisit popular UDP-based protocols of network operators, including DNS, and analyze their behavior in the context of DDoS attacks.</p>	<p>Leveraging Internet Background Radiation for Opportunistic Network Analysis</p> <p>Karyn Benson<sup>1</sup>, Alberto Dainotti<sup>1</sup>, Alex C. Snoeren<sup>1</sup>, Michael Kallitsis<sup>1</sup></p> <p><b>IMC '15</b></p> <p><b>ABSTRACT</b></p> <p>For more than a decade, unsolicited traffic sent to unused regions of the address space has provided valuable insight into malicious Internet activities. In this paper, we explore the utility of this traffic, known as Internet Background Radiation (IBR), for a different purpose: as a data source of Internet-wide measurements. We collect and analyze IBR from two large datasets, carefully deconstructing its various components and characterizing them along dimensions applicable to Internet-wide measurements. Intuitively, IBR is a rich source of information about the address space, known as Internet Background Radiation (IBR), to address this challenge. Monitoring unused portions of the IPv4 address space reveals that IBR is of considerable volume, incessant, and originates from a variety of services [41, 48]. This unsolicited traffic is caused by scanning (e.g., searching for hosts running a vulnerable service), misconfigurations (e.g., a typo in the IP address for a mail server), backscatter (responses to packets with forged source IP addresses, including spoofed DoS attack), bugs, etc. Historically, researchers</p>	<p>Cleaning Up the Internet of Evil Things: Real-World Evidence on ISP and Consumer Efforts to Remove Mirai</p> <p><b>NDSS '19</b></p> <p><b>Abstract</b></p> <p>While it is well known that Mirai is still the leading malware family and responsible for 21% of the IoT infected devices [3], what keeps Mirai a relevant threat is that it exploits default credentials, a problem that has still not been fixed by many manufacturers. The Open Web Application Security Project (OWASP) describes this as the top threat for IoT [4]. Additionally, the release of Mirai's source code has allowed attackers to add exploit code on top of its credential-based attacks and create newer variants which</p>	<p>Can ISPs Help Mitigate IoT Malware? A Longitudinal Study of Broadband ISP Security Efforts</p> <p>Aman Norouzi<sup>1</sup>, Elsa Tuciós Rodríguez<sup>1</sup>, Eimer Lisdridge<sup>1</sup>.</p> <p><b>EURO S&amp;P '21</b></p> <p><b>Abstract</b></p> <p>While it is well known that Mirai is still the leading malware family and responsible for 21% of the IoT infected devices [3], what keeps Mirai a relevant threat is that it exploits default credentials, a problem that has still not been fixed by many manufacturers. The Open Web Application Security Project (OWASP) describes this as the top threat for IoT [4]. Additionally, the release of Mirai's source code has allowed attackers to add exploit code on top of its credential-based attacks and create newer variants which</p>
<p>Estimating Internet Address Space Usage through Passive Measurements</p> <p>Alberto Dainotti, Michael Kallitsis, Eduard Glatz, Karyn Benson</p> <p><b>SIGCOMM '14</b></p> <p><b>ABSTRACT</b></p> <p>One challenge in understanding the evolution of Internet infrastructure is the lack of systematic mechanisms for monitoring the extent to which allocated IP addresses are actually used. Address utilization has been monitored via actively scanning the entire IPv4 address space. We evaluate the potential to leverage passive network traffic measurements in addition to or instead of active probing. Passive traffic measurements introduce no network traffic overhead</p>	<p>IoT POT: Analysing the Rise of IoT Compromises</p> <p>Yin Minn Pa Pa<sup>1</sup>, Shogo Suzuki<sup>1</sup>, Katsunari Yoshioka<sup>1</sup>, Tsutomu Matsumoto<sup>1</sup>, Takahiro Kasama<sup>2</sup>, Christian Rossow<sup>3</sup></p> <p><b>USENIX WOOT '15</b></p> <p><b>Abstract</b></p> <p>We analyze the increasing threats against IoT devices. We show that Telnet-based attacks that target IoT devices have rocketed since 2014. Based on this observation, we propose an IoT honeypot and sandbox, which attracts and analyzes Telnet-based attacks against vari-</p>	<p>Internet-scale Probing of CPS: Inference, Characterization and Orchestration Analysis</p> <p>Claude Fatchka<sup>1,2</sup>, Elias Bou-Harb<sup>3</sup>, Anastasis Keliotis<sup>1</sup>, Nasir Memon<sup>1</sup>, and Mustaque Ahamad<sup>1,4</sup></p> <p><b>NDSS '17</b></p> <p><b>Abstract</b></p> <p>Although the security of Critical Infrastructure Systems (CIS) has been recently receiving significant attention from the research community, undoubtedly, there still exists a substantial lack of a comprehensive and a holistic understanding of attackers' malicious strategies, aims and intentions. To this end, this paper uniquely exploits passive monitoring and analysis of a newly deployed network telescope IP address space in a first attempt ever to build broad notions of real CPS maliciousness. Specifically, we approach this problem by inferring, investigating, characterizing, and correlating from only available activities that</p>	<p>DANTE: A Framework for Mining and Monitoring Darknet Traffic</p> <p>Dvir Cohen, Yusef Minsky, Yoram Elovick, Barak Peiris, and Anaf Shabtai</p> <p><b>ESORICS '20</b></p> <p><b>Abstract</b></p> <p>Darknet traffic is a rich source of information about the Internet. However, obtaining these insights from darknet traffic is a challenging task for three reasons: (1) Darknet IPs are not assigned to actual hosts so the traffic only captures the initiation of a communication, and not the actual content of the traffic; (2) Darknet traffic is often encrypted, which makes it difficult to analyze; (3) Darknet traffic is often fragmented, which makes it difficult to analyze. In this paper, we present DANTE, a framework for mining and monitoring darknet traffic. DANTE consists of three main components: (1) DANTE-Collector, which captures darknet traffic; (2) DANTE-Analyzer, which analyzes darknet traffic; (3) DANTE-Visualizer, which visualizes darknet traffic. DANTE is the first framework to capture darknet traffic at the application level, and to analyze it at the application level. DANTE is the first framework to capture darknet traffic at the application level, and to analyze it at the application level.</p>	



# NICTERデータセットはNONSTOP上で提供

- サイバーセキュリティ情報を遠隔から安全に利用してもらうための環境



# NICTER Dataset まとめ

---

- 今年度提供するデータは2種類：
  - ダークネットトラフィック
  - スпамメールデータ（要望があれば）
- データセットはNONSTOP上で提供：
  - データにアクセスできるVM環境をユーザ毎に用意
  - 利用申請は [nonstop@ml.nict.go.jp](mailto:nonstop@ml.nict.go.jp) まで（継続の方も要連絡）
- メリット：
  - リアルタイムかつ継続的(10年間以上)のデータセット
  - 加工されていない生データなので用途は自由